Discrimination of *pp* solar neutrinos and ^{14}C double pile-up events in a large-scale LS detector[*]

Guo-Ming Chen,[1] Xin Zhang,[2,3] Ze-Yuan Yu,[2] Si-Yuan Zhang,[1]
Yu Xu,[4] Wen-Jie Wu,[5] Yao-Guang Wang,[2] and Yong-Bo Huang[1,†]

[1]*School of Physical Science and Technology, Guangxi University, Nanning 530004, China*
[2]*Institute of High Energy Physics, Beijing 100049, China*
[3]*University of Chinese Academy of Sciences, Beijing 100049, China*
[4]*School of Physics, Sun Yat-Sen University, Guangzhou 510275, China*
[5]*Department of Physics and Astronomy, University of California, Irvine, California, USA*

As a unique probe, the precision measurement of *pp* solar neutrinos is important for studying the sun's energy mechanism as it enables monitoring the thermodynamic equilibrium and studying neutrino oscillations in the vacuum-dominated region. For a large-scale liquid scintillator detector, a bottleneck for *pp* solar neutrino detection is the pile-up events of intrinsic ^{14}C decay. This paper presents a few approaches to discriminating between *pp* solar neutrinos and ^{14}C pile-up events by considering the differences in their time and spatial distributions. In this study, a Geant4-based Monte Carlo simulation is conducted. Multivariate analysis and deep learning technology are adopted to investigate the capability of ^{14}C pile-up reduction. The BDTG model and VGG network demonstrate good performance in discriminating *pp* solar neutrinos and ^{14}C double-pile-up events. Under the ^{14}C concentration assumption of $5 \times 10^{-18} g/g$, the signal significance can achieve 10.3 and 15.6 using the statistics of only one day. In this case, the signal efficiency for discrimination using the BDTG model while rejecting 99.18% ^{14}C double pile-up events is 51.1%, and that for the case where the VGG network is used while rejecting 99.81% of the ^{14}C double pile-up events is 42.7%.

Keywords: Liquid scintillator detector, pp solar neutrinos, ^{14}C pile-up, Multivariate analysis, Deep learning

I. INTRODUCTION

With the development of nuclear physics and astrophysics, we have been able to glimpse into the sun's energy mechanism, which originates from the nuclear fusion of light nuclei in the core of the sun [1–3]. The proton-proton (*pp*) cycle produces ∼99% of solar energy, and its primary reaction is the fusion of two protons into a deuteron:

$$p + p = {}^{2}\text{H} + e^{+} + \nu_e \qquad (1)$$

In this reaction, large numbers of low-energy neutrinos, called *pp* neutrinos, are emitted ($E < 0.42$ MeV). In addition, the proton-electron-proton (*pep*) process and secondary reactions in the *pp* cycle also emit neutrinos known as *pep* neutrinos, ^{7}Be neutrinos, ^{8}B neutrinos, and *hep* (helium-proton) neutrinos. The remaining energy of the sun is contributed by the carbon-nitrogen-oxygen (CNO) cycle, which emits CNO neutrinos. The detection of solar neutrinos is considered a direct way to test theoretical solar models. However, differences between early observations and theoretical predictions were discovered [4–13], leading to the so-called "solar neutrino problem" that has plagued us for more than 30 years. Subsequently, the MSW-LMA mechanism [14, 15] was proven to be the standard solution based on solid evidence provided by SNO [16, 17] and KamLAND [18]. Currently, the standard solar model (SSM) [19–24] provides precise predictions of the flux and energy distribution of solar neutrinos. Almost all solar neutrino components have been observed [25–28], and we expect to enter an era of precise and comprehensive measurements of solar neutrinos in the coming decades [29, 30].

pp neutrinos are strongly related to the predominant energy production of the sun and carry recent messages from the core of the sun. These characteristics make them an important means for studying the sun's energy mechanism and thermodynamic equilibrium monitoring. By contrast, *pp* neutrinos can be used to study neutrino oscillations in vacuum-dominated regions. The detection of *pp* neutrinos simultaneously requires a low threshold (∼ 200keV) along with effective background reduction. *pp* neutrinos were first detected using ^{71}Ga-based radiochemical detectors [6–11]. Subsequently, a large-scale liquid scintillator (LS) detector was successfully applied in a Borexino experiment and provided the best measurement of *pp* neutrinos at the ∼10% level [26, 27] via elastic neutrino-electron scattering.

According to the experience gained from the Borexino experiment, intrinsic ^{14}C decays from an organic liquid scintillator and their associated pile-up events are a crucial internal background for a large-scale LS detector. ^{14}C pile-up events correspond to cases in which more than one ^{14}C decay occurs at different detector positions but in the same trigger window. In addition, pile-ups can be classified into the following categories according to the multiplicity of ^{14}C accidental coincidences: double pile-ups, threefold pile-ups, and fourfold pile-ups. The Borexino experiment (∼278 ton) requires considerable effort for LS purification to obtain a ^{14}C concentration of approximately 2.7×10^{-18}g/g. At this concentration, the ^{14}C double pile-up accounts for approximately 10% of the events

in the spectral gap between the ^{14}C and ^{210}Po spectra [26].

For an LS detector with a sensitive target mass of m kilotons (kton), the frequency of a ^{14}C single event is

$$f_{single}[Hz] = \frac{C_{^{14}C} \cdot N_A \cdot m}{\tau \cdot M} \times 10^9 \qquad (2)$$

where N_A is Avogadro's constants (6.023×10^{23}) and τ, M, $C_{^{14}C}$ correspond to ^{14}C's lifetime, molar mass, and its concentration in the LS, respectively.

The frequency of ^{14}C pile-up events can be calculated as follows:

$$f_{pile-up}[Hz] = \frac{e^{-f_{single} \cdot \Delta t}}{(n-1)!} \cdot f_{single}^n \cdot \Delta t^{n-1} \cdot \varepsilon \quad (3)$$

where n ($n \geq 2$) denotes the multiplicity of the ^{14}C accidental coincidence; for example, $n = 2$ represents the case of a double ^{14}C pile-up. Δt is the time window for detection and ε corresponds to the reconstruction efficiency of the ^{14}C pile-up events.

As the detector mass increases, a dramatic increase in ^{14}C pile-up events must be considered and effectively rejected. Taking a large spherical LS detector as an example, with the radius of the detector being 15 m and the detector mass being approximately 12 kton, Table 1 lists the event rates of pp neutrinos and ^{14}C single and pile-up events at different ^{14}C concentrations. A 500 ns time window was used in this calculation, and the reconstruction efficiency was set to 100%. For a ^{14}C concentration of 5×10^{-18} g/g in the LS of the above detector, Fig. 1 shows the recoil energy spectra of pp neutrinos via elastic neutrino-electron scattering, which can be found in [30]. The energy spectra of ^{14}C single, double, and triple pile-up events are shown for comparison. In this giant detector, the ^{14}C pile-up events completely outnumbered the pp neutrino signals by more than two orders of magnitude.

In Table 1, the values in brackets indicate the event rates within the energy range of interest of 0.16 MeV–0.25 MeV for the deposited energy, considering that the Q value of ^{14}C β decay is $\sim$ 156 keV and the scattered electron of the pp neutrino reaction is difficult to distinguish from the emitted electron of a ^{14}C single event. The target mass of the above detector ($\sim$12 kton) is $\sim$43 times larger than that of Borexino ($\sim$278 ton). Consequently, the signal-to-background ratio of pp neutrinos and ^{14}C double pile-up events in this detector is smaller than 1:126 for a ^{14}C concentration of 2.7×10^{-18} g/g, and the signal-to-background ratio will be much lower if a higher ^{14}C concentration is used. However, because the energy resolution introduces smearing in the energy spectrum, the energy range of the analysis must be determined based on realistic situations.

More neutrino experiments are underway or are being planned, and many of them [31–35] have good potential for pp neutrino detection because they are expected to have a large detector target, well-controlled radioactivity, low detection threshold, or good energy resolution. In experiments
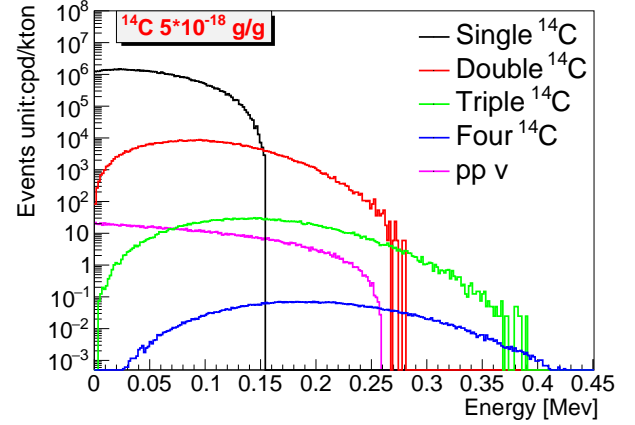
Fig. 1. The recoil energy spectra of pp neutrinos, ^{14}C single, double, and triple pile-up events in a spherical LS detector, whose radius and ^{14}C concentration are 15 m and 5×10^{-18} g/g, respectively. The spectra do not include the detection effects: energy non-linearity, non-uniformity, and resolution. The higher order contribution from the ^{14}C pile-up is negligible and not shown.

with LS detectors of the order of tens of kilotons, neutrino detection in low-energy regions is difficult because of ^{14}C pile-up. Therefore, an approach must be developed for ^{14}C pile-up discrimination and reduction, especially that for ^{14}C double pile-up, because its event rate is much higher than that of other accidental coincidences.

This study focuses on discriminating between pp solar neutrinos and ^{14}C double pile-up events. The discrimination of other accidental coincidences with a ^{14}C multiplicity ≥ 3 is an important topic in the case of a higher ^{14}C concentration; however, it is not the subject of this study. The details of our work are as follows: First, we simulated an LS detector and investigated the features of the detector's PMT hit pattern for pp neutrinos and ^{14}C double pile-up events (Sec. II). We then present several approaches to ^{14}C double-pile-up discrimination based on multivariate analysis and deep-learning technology (Sec. III). In Sec. IV, the discrimination performances are shown and compared. Finally, a summary is presented in Sec. V

II. DETECTOR SIMULATION

In this study, a spherical LS detector was built using Monte Carlo (MC) simulations with the Geant4 toolkit [36] version 4.10.p02. The radius of the spherical detector was 15 m, and the LS was contained in an acrylic sphere with a 10-cm-thick wall. To simplify the simulation, a sensitive optical surface was defined for receiving the photons instead of using the detailed PMT simulation. The sensitive optical surface was a sphere outside the acrylic sphere, separated by a 1-m-thick layer of water. The coverage and quantum efficiency of the photosensors could be easily tuned. In the simulation, the coverage rate was 65%, corresponding to ap-

Table 1. The event rates (unit: cpd/kton) of *pp* neutrinos and ^{14}C single and pile-up events in different ^{14}C concentrations. A spherical LS detector ($\sim$12 kton) with a 15-m radius was used in the calculation, and the time window was 500 ns. The values in the brackets indicate the event rates inside the energy range of interest (0.16, 0.25) MeV; the ratio is about 10% for both *pp* neutrinos and ^{14}C double pile-up events.

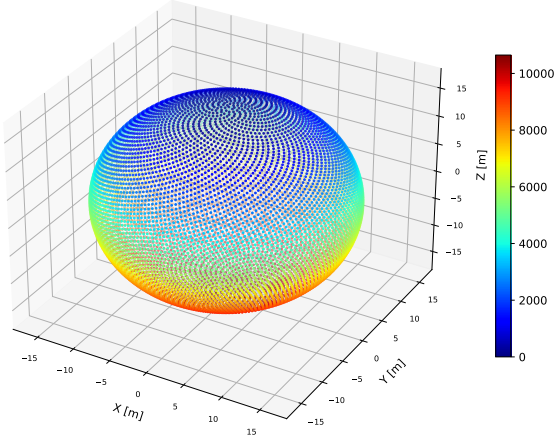| Event types	$10^{-18}g/g$	$2.7 \times 10^{-18}g/g$ (Borexino-like)	$5 \times 10^{-18}g/g$	$10^{-17}g/g$
pp-ν	1.37×10^3 ($\sim 1.37 \times 10^2$)	1.37×10^3 ($\sim 1.37 \times 10^2$)	1.37×10^3 ($\sim 1.37 \times 10^2$)	1.37×10^3 ($\sim 1.37 \times 10^2$)
^{14}C single	1.43×10^7	3.86×10^7	7.16×10^7	1.43×10^8
^{14}C double	2.38×10^4 ($\sim 2.38 \times 10^3$)	1.73×10^5 ($\sim 1.73 \times 10^4$)	5.94×10^5 ($\sim 5.94 \times 10^4$)	2.38×10^6 ($\sim 2.38 \times 10^5$)
^{14}C triple	1.97×10^1	3.88×10^2	2.47×10^3	1.97×10^4
Signal-to-background ratio: ($\frac{pp-\nu}{^{14}Cdouble}$)	$\sim 1 : 17$	$\sim 1 : 126$	$\sim 1 : 431$	$\sim 1: 1727$

Fig. 2. A schematic view of the detector. Each pixel corresponds to a 20-inch PMT, and its color indicates the ID of each PMT. In total, the detector had 10650 PMTs.

Table 2. PMT parameters in the simulation.

Parameters	Values
PMT Coverage	65%
PMT QE	30% $\pm$ 2% (Gaussian)
PMT TTS	3 $\pm$ 0.3 ns (Gaussian)
PMT dark rate (DR)	20 $\pm$ 3 kHz (Gaussian)
PMT spe resolution	30% $\pm$ 3% (Gaussian)
Time window	500 ns

proximately 10650 twenty-inch photomultipliers (PMTs) uniformly distributed on the sensitive optical surface. Fig. 2 shows a schematic of the detector. In the simulation, an average quantum efficiency of 30% was used for the 20-inch PMTs with a 2% Gaussian relative spread. The LS properties were referenced from [37–44], and comprehensive optical processes were adopted, including quenching, Rayleigh scattering, absorption, reemission, photon transport in the LS, and reflections on the acrylic surface. Table 2 summarizes the main parameters of the PMTs in the simulation, including the transit time spread (TTS), quantum efficiency (QE), dark noise (DR), and resolution of a single photoelectron (spe). As a result, approximately 1100 photoelectrons (PEs) could be observed by the 10650 PMTs for a 1 MeV electron that fully deposited its kinetic energy in the center of the detector, which corresponds to approximately a 3% energy resolution. In contrast, approximately 106.5 additional PEs originating

from the PMT dark noise in a time window of 500 ns could be detected.

To investigate the response features of *pp* neutrinos and ^{14}C double-pile-up events, MC samples were generated and compared. Approximately one million final-state electrons from the elastic neutrino-electron scattering reaction of *pp* neutrinos were uniformly simulated in the LS volume, and the spectrum of the scattered electrons was referenced from [30]. Because the final-state electrons from elastic neutrino-electron scattering are similar to the electrons emitted from the ^{14}C β decay (^{14}C single event), distinguishing them at an event-by-event level is difficult. Therefore, an energy reduction is required to focus on a narrow energy region. The same treatment method as used by Borexino et al. was used. However, electrons whose kinetic energy is approximately 200 keV in LS show a 5% energy nonlinearity [44, 45], and the energy resolution is already included in the above simulation. As a result, in our analysis, a 270 PE cut was applied to the total number of PEs of all PMTs by considering the $\sim$156 keV end-point energy of the ^{14}C β decay ($\sim$163 PEs) and the contribution of PMT dark noise ($\sim$106.5 PEs).

After the total PE cut, an MC sample that included 100,000 *pp* neutrinos was used for the discrimination study, and they were uniformly distributed in the LS. To generate the ^{14}C double pile-up sample, first a large dataset was produced by simulating 10 million ^{14}C single events in the LS using ^{14}C
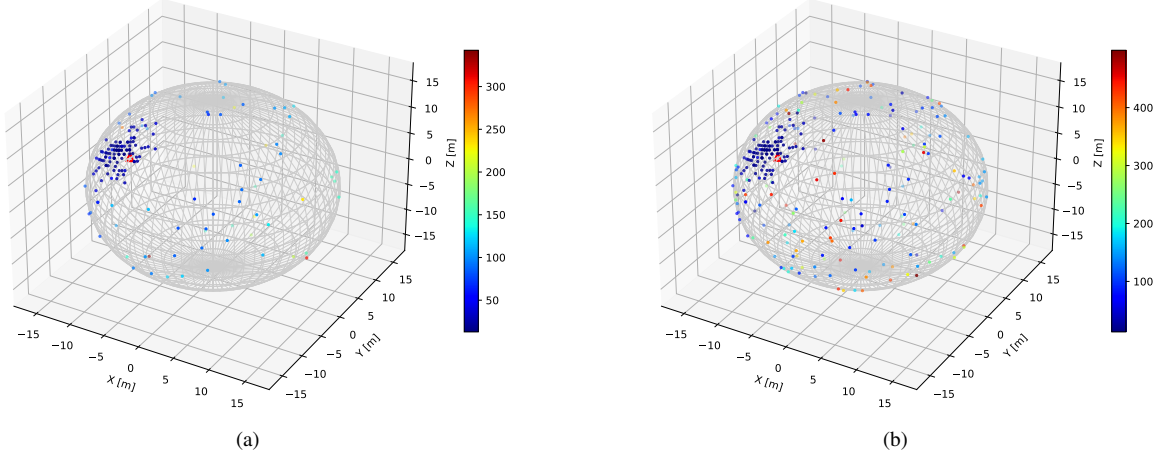
Fig. 3. The PMT hit patterns of a *pp* solar neutrino event. Each pixel corresponds to a fired PMT, and its color indicates the hit time information. The location of the red hollow triangle is (-6582.21, -8972.86, 8696.34) mm, which indicates the position where the physical event deposited its energy (159.94 keV). (a) only physical hits are included, and 172 PEs are observed for a 500-ns time window. (b) Both physical hits and PMT dark noise hits are shown, and 284 PEs are observed for a 500-ns time window, including 112 PEs from the PMT dark noise.
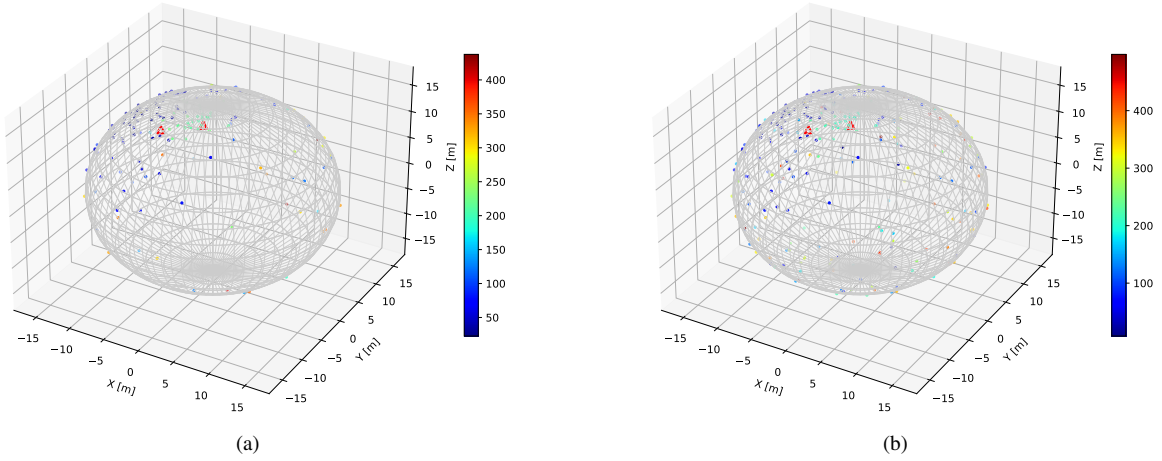

Fig. 4. The PMT hit patterns of a ^{14}C double pile-up event. Each pixel corresponds to a fired PMT and its color indicates the hit time information. The two red hollow triangles indicate the positions where two ^{14}C events deposited their energies (71.161 keV and 56.593 keV). Their locations are (-6229.32, -2139.36, 10471.7) mm and (484.61, -3199.44, 14423.5) mm, respectively. (a) only physical hits are included, and 173 PEs (107+66) are observed for a 500-ns time window. (b) Both physical hits and PMT dark noise hits are shown, and 273 PEs are observed for a 500-ns time window, including 100 PEs from the PMT dark noise.

β decay. Next, two ^{14}C single events were randomly selected from the dataset and merged into a double-pile-up event. In the merge operation, because the lifetime of ^{14}C is longer than 8000 years, the time interval between two ^{14}C single events could be considered an approximately uniform distribution for a few hundred nanoseconds. Similarly, a 270 PE cut was applied, and 100,000 ^{14}C double pile-up events were used for our analysis.

As illustrated in Figs. 3 and 4, *pp* solar neutrinos and ^{14}C double pile-up events exhibited different features in their temporal and spatial distributions. The *pp* solar neutrino is a

single point-like event whose energy deposition occurs in a relatively short time and small space; hence, only one cluster is expected to be found in its PMT hit pattern. For the ^{14}C double-pile-up event, if two ^{14}C atoms decay at different detector positions, two clusters are expected to be found. However, because the hit time distribution of the fired PMTs includes both the scintillation time and the photon's time of flight, as well as the decay time of ^{14}C, the hit time distribution is useful for identification studies. In particular, when two ^{14}C atoms decay near each other, their spatial distribution is not expected to be significantly different from that of

a single point-like event. However, the hit time distribution is still helpful if the time interval between the two ^{14}C decays is large. An example is shown in Fig. 4.

As mentioned previously, our approach employs a straightforward trigger strategy that considers whether the total number of PEs within 500 ns exceeds 270 PEs. Subsequently, we selected the hit information within this timeframe for further analysis. However, the trigger strategy must be optimized. As described in Sec. III, the event spatial information was extracted and used together with the hit time information as input to the discrimination algorithms.

III. DISCRIMINATION METHODS

The basic idea behind developing a discrimination algorithm for *pp* solar neutrinos and ^{14}C double pile-up events is to utilize their temporal and spatial information, which have different characteristics (see Sec. II). Similar approaches have been applied to discriminate single-site and multisite energy depositions in large-scale liquid scintillation detectors [46]. During the measurement, the cluster structure was smeared by interference from dark noise and the TTS of the PMT. These effects make the identification more challenging and more efficient approaches are required. In this study, a multivariate analysis using the Toolkit for Multivariate Data Analysis (TMVA) [47, 48] was performed, and the widely used algorithm, boosted decision trees with gradient boosting (BDTG), was chosen for the analysis. In addition, deep learning technologies based on the VGG network were also applied. In the following section, we present details of the discrimination method.

A. TMVA analysis

TMVA [47, 48] is a powerful tool for multivariate analysis and has been successfully applied to both signal and background classification in accelerator physics [49], component identification of cosmic rays [50], and event reconstruction in LS detectors for neutrino experiments [55]. The TMVA toolkit hosts a wide variety of multivariate classification algorithms. In this study, we used the TMVA algorithm, BDTG. To extract the input variables, the PMT hit pattern was projected onto a one-dimensional (1-D) plane for the hit time, θ, and ϕ of each fired PMT in spherical coordinates. The projection results of Fig. 3(b) are shown in Fig. 5, and the projection results of Fig. 4(b) are shown in Fig. 6. The *pp* solar neutrino, which is a single-point-like event, showed only one cluster in its distribution, whereas the ^{14}C double pile-up event showed two clusters.

These 1-D distributions were used in multivariate analysis. The input variables for the TMVA algorithms should be sensitive to discrimination and contain the characteristics of *pp* solar neutrinos and ^{14}C double-pile-up events. In our analysis, we found that hit time information dominated the discrimination performance; therefore, more variables were extracted from the 1-D distribution of hit time. Fifteen variables were

Table 3. Input variables for multivariate analysis.

Variable	Description
$V_1^{hittime}$	Number of hits in the first 200 ns
$V_2^{hittime}$	The peak position of the highest bin in the first 200 ns
$V_3^{hittime}$	The amplitude of the highest bin in the first 200 ns
$V_4^{hittime}$	Number of hits in (200, 500) ns
$V_5^{hittime}$	The amplitude of the highest bin in (200, 500) ns
$V_6^{hittime}$	The ratio between the peak amplitude and the peak position of the highest bin in (200, 500) ns
$V_7^{hittime}$	The ratio between the number of hits in the first 200 ns and in (200, 500) ns
$V_8^{hittime}$	The RMS value of the 1-D distribution of hit time
$V_9^{hittime}$	The Mean value of the 1-D distribution of hit time
V_1^{theta}	The RMS value of the 1-D distribution of θ
V_2^{theta}	The skewness coefficient of the 1-D distribution of θ
V_3^{theta}	The kurtosis coefficient of the 1-D distribution of θ
V_1^{phi}	The RMS value of the 1-D distribution of ϕ
V_2^{phi}	The skewness coefficient of the 1-D distribution of ϕ
V_3^{phi}	The kurtosis coefficient of the 1-D distribution of ϕ

used in the TMVA analysis. These variables are denoted as V_i^α, where $i = 1, 2, 3$, etc., and correspond to the extracted parameters in each 1-D distribution. $\alpha = hittime$, θ, or ϕ, which indicates that the variables are from the 1-D distribution of hit time, θ, or ϕ. The details can be found in Table 3.

Fig. 7 shows the normalized distributions of these input variables, and the difference in their shapes is determined by comparing the two types of events. By contrast, the correlations of the input variables were checked for both *pp* solar neutrinos and ^{14}C double pile-up events. As shown in Fig. 8, because we dropped several variables with strong correlations in a previous study, the correlation of the current variables is acceptable, provided it is less than 90%. A few variables had close to 90% correlations, and we retained them in the analysis. This is mainly determined by considering that the variables exhibit different correlations for the signal and background; a similar strategy was applied in [56].

The MC samples of the *pp* solar neutrinos and ^{14}C double pile-up events were divided into two equal parts, one for TMVA training and the other for validation. Hence, both the training and test samples include 50 thousand pp neutrinos
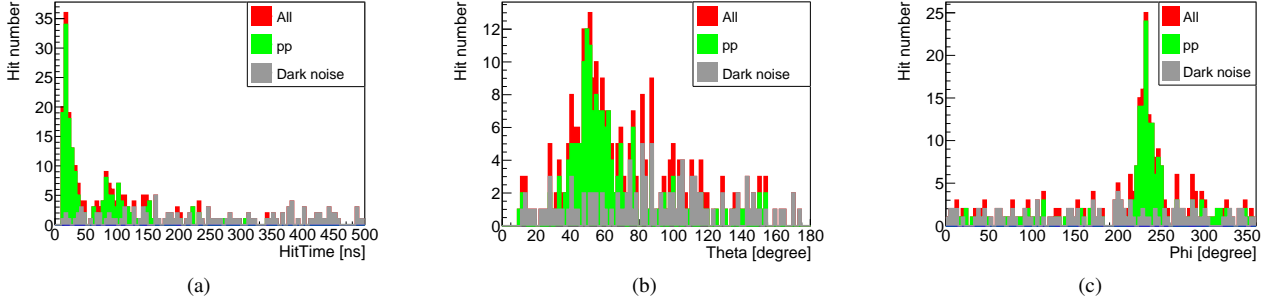
Fig. 5. The hit time, θ, and ϕ distributions of a *pp* solar neutrino event corresponding to the event in Fig. 3(b). (a) Hit time distribution. (b) θ distribution. (c) ϕ distribution.
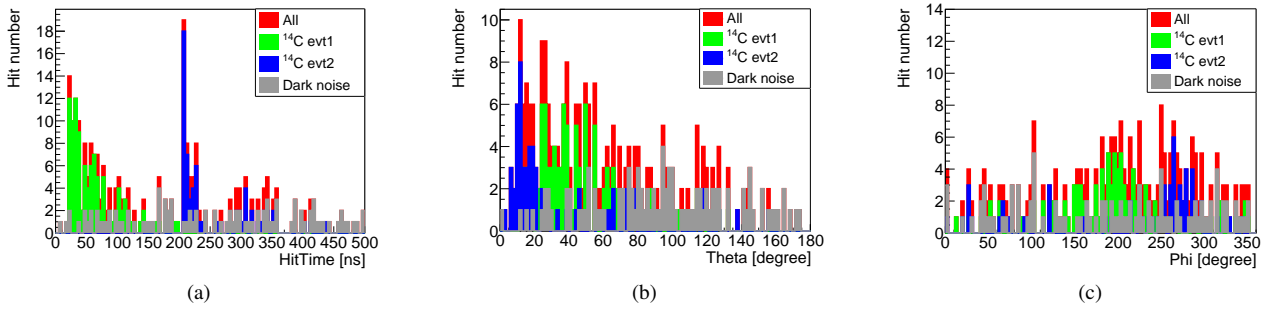

Fig. 6. The hit time, θ, and ϕ distributions of a ^{14}C double pile-up event corresponding to the event in Fig. 4(b). (a) Hit time distribution. (b) θ distribution. (c) ϕ distribution.
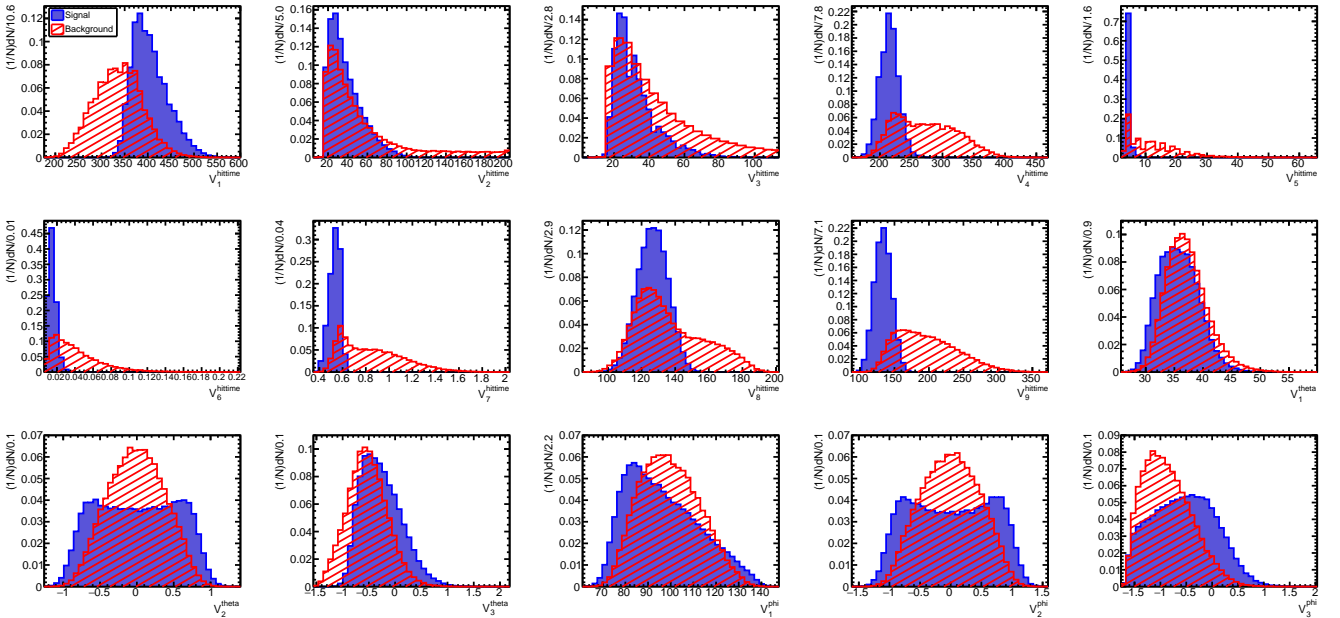

Fig. 7. Normalized distributions of the variables of *pp* solar neutrino and ^{14}C double pileup event.

and 50 thousand C-14 double pile-up events. To improve the performance, several main parameters of the BDTG algorithm were tuned. Table 4 shows the settings of these parameters. The other parameters were set to their default values and are not listed in the tables.
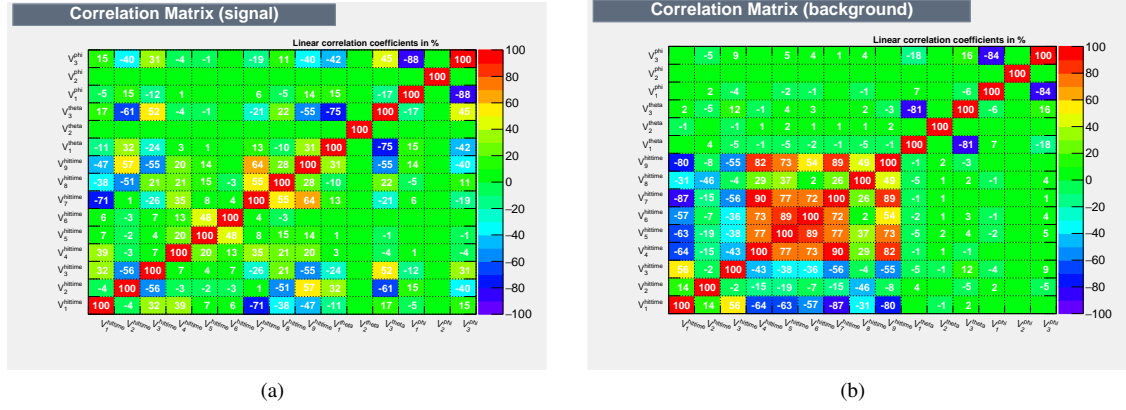
Fig. 8. Linear correlation matrix for the input variables of *pp* solar neutrinos (a) and ^{14}C double pile-up events (b).

Table 4. Parameters used in the BDTG algorithm.

Configuration option	Setting	Description
NTrees	1000	Number of trees in the forest
MaxDepth	2	Max depth of the decision tree allowed
MinNodeSize	2.5%	Minimum percentage of training events required in a leaf node
nCuts	20	Number of grid points in variable range used in finding optimal cut in node splitting
BoostType	Grad	Boosting type for the trees in the forest

B. Deep learning

Deep learning technology is widely used in high-energy and nuclear physics, with many successful applications [51–55, 57, 58] such as energy reconstruction, track reconstruction, particle identification, and signal processing. In this study, the deep learning algorithm VGG convolutional neural network was used for the feature recognition of one-dimensional sequences. The extracted PMT hit patterns are projected onto a one-dimensional feature series for the hit time, θ, and ϕ, respectively. The resulting patterns are similar to those in Figs. 5 and 6. To extract their features, a one-dimensional convolution kernel was used for the three series, a pooling layer was used for information compression, and a fully connected layer was used for particle classification. The model structure was based on the architecture of VGG-16, which includes 13 convolution and pooling modules, three fully connected layers, batch normalization layers, and connected neural unit dropout processing.

In addition to one-dimensional projection using PMT hit patterns, we also attempted two-dimensional projection methods to provide input to deep learning networks, including the Mercator projection, sinusoidal projection, and a projection method based on the arrangement of PMTs [55, 59]. However, after applying the two-dimensional projection, the performance did not improve but, in fact, slightly worsened. Considering that the number of hits in the energy range of interest is very small, we performed a detailed investigation and comparison because the cluster features were much more pronounced in the one-dimensional projection but very discrete in the two-dimensional projection.

Finally, a one-dimensional projection was used to provide input to the VGG network described above. We trained the VGG network using adaptive momentum with a batch size of 256 samples, momentum of 0.9, and an initial learning rate of 0.01. After every 10 epochs, the learning rate was reduced by a factor of 10. The accuracy of the model was evaluated using a cross-entropy loss function. In the discrimination study using the VGG network, 80% of the *pp* neutrino and ^{14}C double-pile-up separately were used separately for training, whereas the other 20% were used for validation.

IV. DISCRIMINATION PERFORMANCE AND DISCUSSION

A. Discrimination performance of the BDTG model

Fig. 9 shows the training results of the BDTG model. The network was not overtrained, as the responses of the testing data were consistent with those of the training data (Fig. 9(a)). The signal and background were separated into two parts after training; however, some overlapping components remained, indicating that their event features were similar. Hence, the network failed to distinguish between them. According to a detailed investigation, one of the main reasons for the failed identification was the stacking of two ^{14}C atoms that are very close together in both time and space. To op-
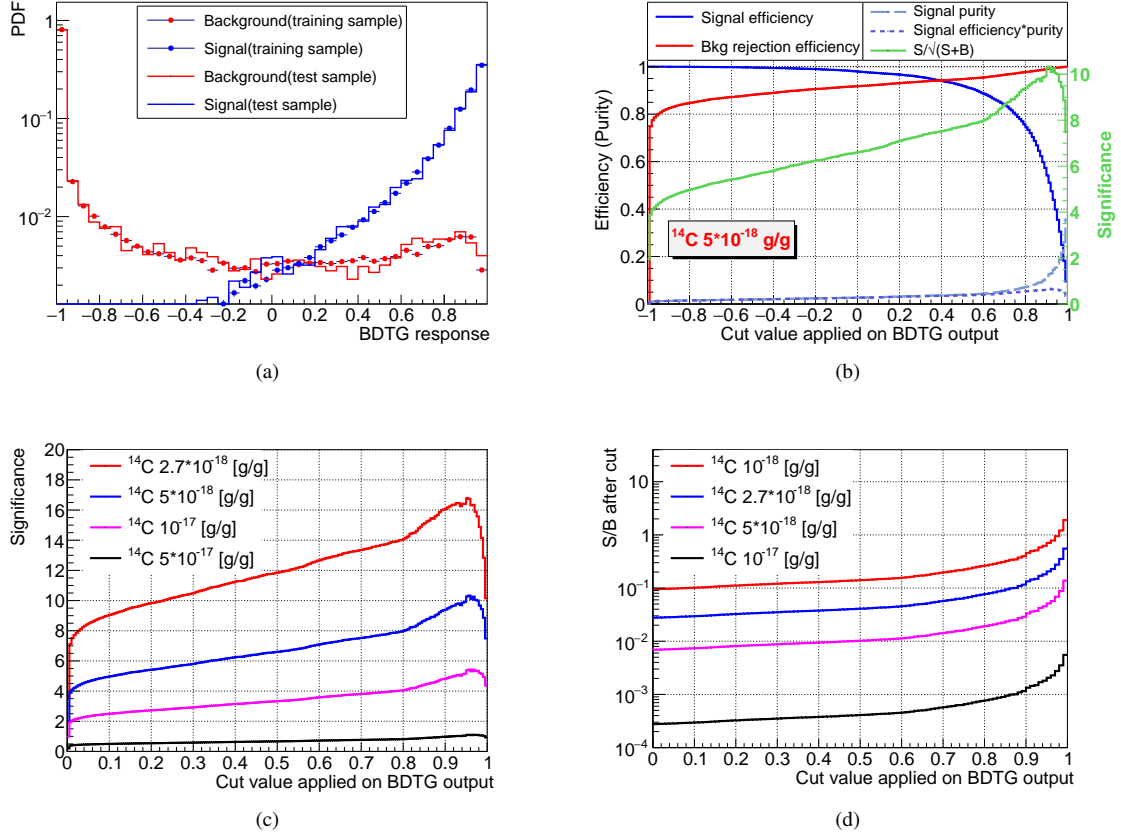
(a)

(b)

(c)

(d)

Fig. 9. Identification performance using the BDTG model. (a) Normalized response distributions of the BDTG model for the signal and the background. (b) Cut efficiencies as functions of BDTG cut values. The significance (green line) was calculated using the statistics for one day of the signal and the background in the analysis region, and the ^{14}C concentration of LS was assumed to be 5×10^{-18} g/g. (c) Significance of different assumptions of the ^{14}C concentration. (d) Signal-to-background ratio after identification for different assumptions of the ^{14}C concentration; the statistics for one day were adopted.

timize the significance $N_s/\sqrt{N_s + N_b}$, where N_s and N_b are the numbers of signals and background after identification, we scanned the cut value on the BDTG response, and the corresponding efficiencies were also obtained. The ^{14}C concentration in the LS was assumed to be 5×10^{-18}g/g, as shown in Fig. 9(b). The significance calculations using the statistics in the analysis region for a period of one day (true energy:160-250 keV), based on the estimation in Table 1, are $\sim$1653 for the signal and $\sim$712440 for the background (considering only the ^{14}C double pile-up events) before the identification. For the BDTG model, the significance reached its maximum value of 10.33 after applying a cut at 0.915, and the signal and background rejection efficiencies were 51.1% and 99.18%, respectively. As discussed in Sec. I, the signal-to-background ratio of the pp neutrinos and ^{14}C double-pile-up events was low in a large-scale LS detector. Therefore, a strict cut is required to reject most of the background. In this case, 51.1% is an acceptable value for signal efficiency, and it still corresponds to a much larger number of effective pp neutrino signals per day compared with most existing experiments.

In Fig. 9(c), the significance was evaluated for different assumptions for ^{14}C concentration. Fig. 9(d) shows

the signal-to-background ratio after identification using the BDTG model, based on the statistics for a period of one day for different ^{14}C concentrations. As a result, the BDTG model exhibits excellent performance and can handle most of ^{14}C double pile-up events effectively.

In addition, other TMVA algorithms were investigated, including the likelihood algorithm and several BDT models (BDT and BDTD). Many exhibited similar performances (Fig. 10), indicating the robustness and stability of the analysis.

B. Discrimination performance of the VGG network

Fig. 11 shows the training results of the VGG network. The network was not overtrained, as the responses of the testing data were consistent with those of the training data (Fig. 11(a)). To optimize the significance, we scanned the cut values of the VGG output, and the corresponding efficiencies were obtained. The ^{14}C concentration of the LS was assumed to be 5×10^{-18}g/g, as shown in Fig. 11(b), and the calculation of significance using the statistics for a period of one day
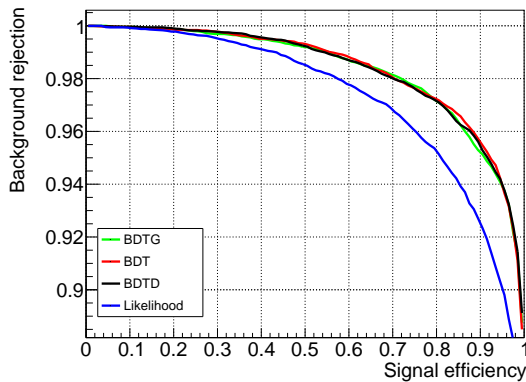
Fig. 10. Relationship between background rejection efficiency and signal efficiency obtained using various TMVA algorithms.

in the analysis region was based on the estimation in Table 1. For the VGG network, the significance reached its maximum value of 15.55 after applying a cut of 0.975. The signal efficiency and background rejection efficiency were 42.7% and 99.81%, respectively.

In Fig. 11(c), the significance was evaluated using different assumptions for ^{14}C concentration, whereas Fig. 11(d) shows the signal-to-background ratio after identification using the VGG network. The calculations were based on the statistics for one day for different ^{14}C concentrations. As a result, the VGG network showed excellent performance and could achieve higher significance and a good improvement in the signal-to-background ratio compared with the BDTG model.

Furthermore, the discrimination performances of the different MC samples were compared, as shown in Fig. 12.

They worsened after including the PMT dark noise, whereas TTS had only a small influence. In addition, the discrimination performance based on the VGG network was stable when rejecting ~99.8% of the ^{14}C double pile-up events.

V. SUMMARY

Large-scale LS detectors have the benefits of a large target mass and high energy resolution, which gives them good potential for *pp* solar neutrino detection. However, they also face a serious problem of the ^{14}C pile-up background. In this study, we investigated how *pp* solar neutrinos and ^{14}C double-pile-up events in a large-scale LS detector could be distinguished using multivariate analysis and deep learning technology. In the simulation study, a spherical LS detector was built using the Geant4 toolkit, and comprehensive optical processes were adopted. The response features in the PMT hit patterns of *pp* neutrinos and ^{14}C double pile-up events were compared, and clear differences were found in their temporal and spatial distributions because one of them was a single point-like event, whereas the other was an accidental coincidence of multiple events.

Using the BDTG model for the *pp* neutrino and ^{14}C double pile-up event discrimination, at a ^{14}C concentration of $5 \times 10^{-18} g/g$, a signal significance of 10.3 could be achieved using the statistics for a period of only one day. The signal efficiency was 51.1% when 99.18% of ^{14}C double-pile-up events were rejected. In the VGG network model, signal significance could reach 15.6 using the statistics for a period of only one day, and the signal efficiency was 42.7% when 99.81% of ^{14}C double pile-up events were rejected. This analysis provides a reliable reference for similar experiments in low-threshold physics detection and ^{14}C pile-up background reduction.

[1] H. A. Bethe, C. L. Critchfield, The formation of deuterium by proton combination. Phys. Rev. **54**, 248-254 (1938). https://doi.org/10.1103/PhysRev.54.248

[2] H.A.Bethe: Energy production in stars. Phys. Rev. 55, 434–456 (1939). https://doi.org/10.1103/PhysRev.55.434

[3] J. N. Bahcall, M.Fukugita and P. I. Krastev, How does the sun shine? Phys. Lett. B **374**, 1-6 (1996). https://doi.org/10.1016/0370-2693(96)00187-6

[4] R. Davis, Jr., D. S. Harmer *et al.*, Search for neutrinos from the sun. Phys. Rev. Lett. **20**, 1205-1209 (1968). https://doi.org/10.1103/PhysRevLett.20.1205

[5] B. T. Cleveland, T. Daily, R. Davis *et al.*, Measurement of the solar electron neutrino flux with a Homestake chlorine detector. Astrophys. J. **496**, 505-526 (1998). https://doi.org/10.1086/305343

[6] P. Anselmann, W. Hampel, G. Heusser *et al.*, Solar neutrinos observed by GALLEX at Gran Sasso. Phys. Lett. B **285**, 376-389 (1992). https://doi.org/10.1016/0370-2693(92)91521-A

[7] W. Hampel, J. Handt, G. Heusser *et al.*, GALLEX Solar Neutrino Observations: Results for GALLEX IV. Phys. Lett. B **447**, 127-133 (1999). https://doi.org/10.1016/S0370-2693(98)01579-2

[8] F. Kaether, W. Hampel, G. Heusser *et al.*, Reanalysis of GALLEX solar neutrino flux and source experiments. Phys. Lett. B **685**, 47-54 (2010). https://doi.org/10.1016/j.physletb.2010.01.030

[9] M. Altmann, M. Balata, P. Belli *et al.*, Complete results for five years of GNO solar neutrino observations. Phys. Lett. B **616**, 174-190 (2005). https://doi.org/10.1016/j.physletb.2005.04.068

[10] J. N. Abdurashitov, V. N. Gavrin, V. V. Gorbachev, *et al.* Measurement of solar neutrino capture rate with gallium metal. III: Results for the 2002–2007 data-taking period. Phys. Rev. C **80**, 015807 (2009). https://doi.org/10.1103/PhysRevC.80.015807

[11] V. N. Gavrin, The history, present, and future of SAGE (et-American gallium experiment). https://doi.org/10.1142/9789811204296_0002

[12] K. S. Hirata, T. Kajita, T. Kifune *et al.*, Observation of B-8 Solar Neutrinos in the Kamiokande-II Detector. Phys. Rev. Lett. **63**, 16 (1989). https://doi.org/10.1103/PhysRevLett.63.16

[13] Y. Fukuda, T. Hayakawa, K. Inoue *et al.*, Solar neutrino data covering the solar cycle 22, Phys. Rev. Lett. **77**, 1683-1686
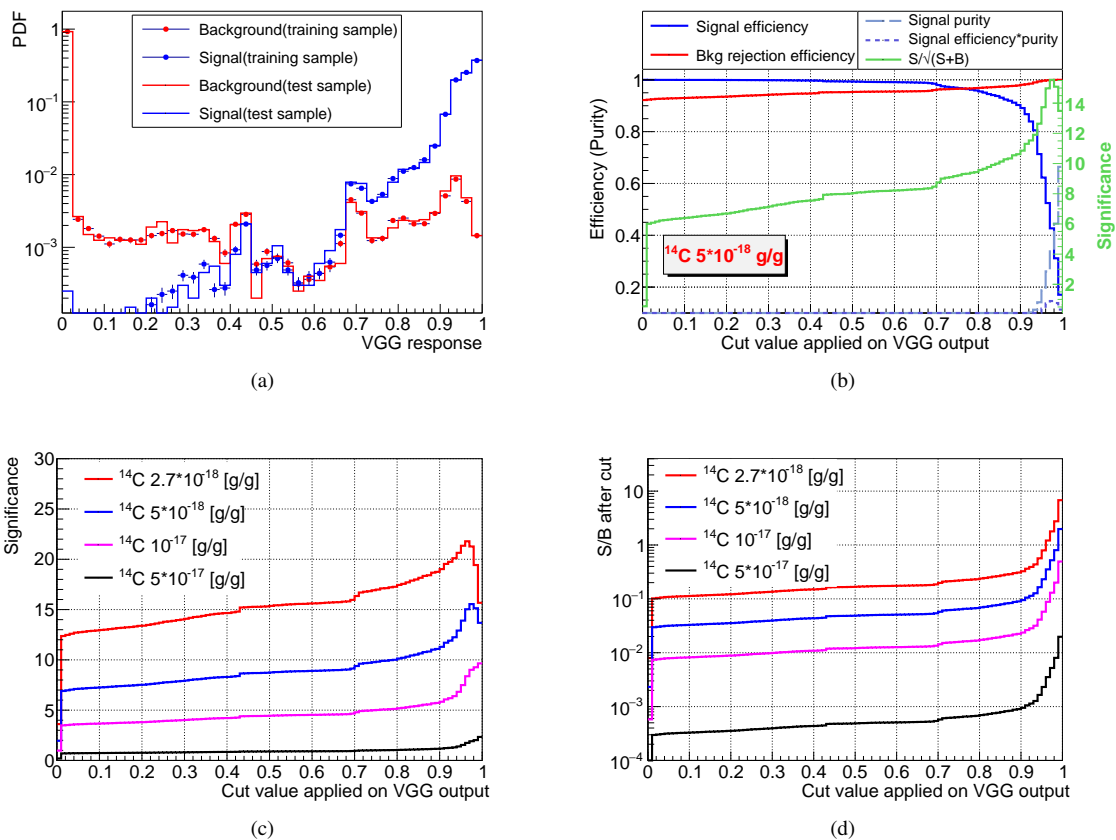
Fig. 11. Identification performance using the VGG network. (a) Normalized response distributions of the VGG network for the signal and the background. (b) Cut efficiencies as functions of VGG cut values. The significance (green line) was calculated using the statistics of the signal and the background in the analysis region for a one-day period; the ^{14}C concentration in LS was assumed to be 5×10^{-18} g/g. (c) Significance for different assumptions of ^{14}C concentration. (d) Signal-to-background ratio after identification for different assumptions of the ^{14}C concentration, the statistics for a one-day period.
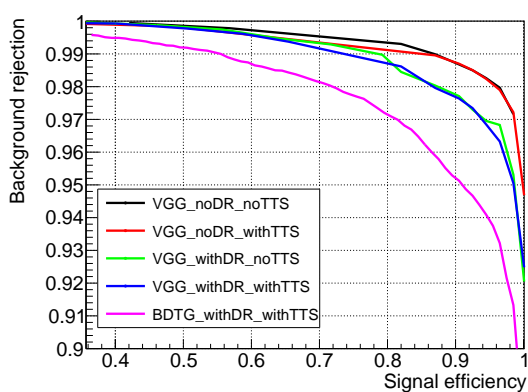

Fig. 12. Relationship between the background rejection efficiency and the signal efficiency for different MC samples.

(1996). https://doi.org/10.1103/PhysRevLett.77.1683

[14] L. Wolfenstein, Neutrino Oscillations in Matter. Phys. Rev. D **17**, 2369-2374 (1978). https://doi.org/10.1103/PhysRevD.17.2369

[15] S. P. Mikheyev, A. Y. Smirnov, Resonance amplification of oscillations in matter, Spectroscopy of solar neutrinos. Sov. J. Nucl. Phys. **42**, 913-917 (1985).

[16] Q. R. Ahmad, R. C. Allen, J. D. Anglin *et al.*, Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory, Phys. Rev. Lett. **87**, 071301 (2001). https://doi.org/10.1103/PhysRevLett.87.071301

[17] S. N. Ahmed, A. E. Anthony, E. W. Beier *et al.*, Measurement of the total active B-8 solar neutrino flux at the Sudbury Neutrino Observatory with enhanced neutral-current sensitivity. Phys. Rev. Lett. **92**, 181301 (2004). https://doi.org/10.1103/PhysRevLett.92.181301

[18] K. Eguchi, S. Enomoto, K. Furuno *et al.*, First results from KamLAND: Evidence for reactor anti-neutrino disappearance. Phys. Rev. Lett. **90**, 021802 (2003). https://doi.org/10.1103/PhysRevLett.90.021802

[19] J. N. Bahcall, M. H. Pinsonneault, Solar models with helium and heavy element diffusion. Rev. Mod. Phys. **67**, 781-808 (1995). https://doi.org/10.1103/RevModPhys.67.781

[20] J. Christensen-Dalsgaard, W. Dappen, S. V. Ajukov *et al.*, Current state of solar modeling. Science **272**, 1286-1292 (1996). https://doi.org/10.1126/science.272.5266.1286

[21] S. Degl'Innocenti, W.A. Dziembowski, G. Fiorentini *et al.*, Helioseismology and standard solar models. Astropart. Phys. **7**, 77-95 (1997). https://doi.org/10.1016/S0927-6505(97)00004-2

[22] A. S. Brun, S. Turck-Chieze and J. P. Zahn, Standard solar models in light of new helioseismic constraints. 2. Mixing below convective zone. Astrophys. J. **525**, 1032-1041 (1999). https://doi.org/10.1086/307932

[23] J. N. Bahcall, The Luminosity constraint on solar neutrino fluxes. Phys. Rev. C **65**, 025801 (2002). https://doi.org/10.1103/PhysRevC.65.025801

[24] A. Serenelli, S. Basu, J. W. Ferguson *et al.*, New solar composition: the problem with revised solar models. Astrophys. J. Lett. **705**, L123-L127 (2009). https://doi.org/10.1088/0004-637X/705/2/L123

[25] C. Arpesella, G. Bellini, J. Benziger *et al.*, First real-time detection of Be-7 solar neutrinos by Borexino. Phys. Lett. B **658**, 101-108 (2008). https://doi.org/10.1016/j.physletb.2007.09.054

[26] G. Bellini, J. Benziger, D. Bick *et al.*, Neutrinos from the primary proton–proton fusion process in the sun. Nature **512**, no.7515, 383-386 (2014). https://doi.org/10.1038/nature13702

[27] M. Agostini, K. Altenmüller, S. Appel *et al.*, Comprehensive Measurement of *pp*-Chain Solar Neutrinos. Nature **562**, no.7728, 505-510 (2018). https://doi.org/10.1038/s41586-018-0624-y

[28] M. Agostini, K. Altenmüller, S. Appel *et al.*, Experimental evidence of neutrinos produced in the CNO fusion cycle in the sun. Nature **587**, 577-582 (2020). https://doi.org/10.1038/s41586-020-2934-0

[29] G. D. O. Gann, K. Zuber, D. Bemmerer *et al.*, The Future of Solar Neutrinos. Ann. Rev. Nucl. Part. Sci. **71**, 491-528 (2021). https://doi.org/10.1146/annurev-nucl-011921-061243

[30] X. J. Xu, Z. Wang, S. Chen, Solar neutrino physics. https://doi.org/10.48550/arXiv.2209.14832

[31] Abusleme *et al.*, JUNO Physics, and Detector. Prog. Part. Nucl. Phys. **123**, 103927 (2022). https://doi.org/10.1016/j.ppnp.2021.103927

[32] J. F. Beacom, S. M. Chen, J. P. Cheng *et al.*, Physical prospects of Jinping neutrino experiment. Chin. Phys. C **41**, no.2, 023002 (2017). https://doi.org/10.1088/1674-1137/41/2/023002

[33] J. Aalbers, F. Agostini, S. E. M. Ahmed Maouloud *et al.*, Solar neutrino detection sensitivity in DARWIN via electron scattering. Eur. Phys. J. C **80**, no.12, 1133 (2020). https://doi.org/10.1140/epjc/s10052-020-08602-7

[34] L. Bieger, T. Birkenfeld, D. Blum *et al.*, Potential for precision measurement of solar PP neutrinos in the Serappis experiment. Eur. Phys. J. C **82**, no.9, 779 (2022). https://doi.org/10.1140/epjc/s10052-022-10725-y

[35] F.P. An, G.P An, Q. *et al.* and neutrino physics using JUNO. J. Phys. G **43**, no.3, 030401 (2016). https://doi.org/10.1088/0954-3899/43/3/030401

[36] S. Agostinelli, J. Allison, K. Amako *et al.*, GEANT4–a simulation toolkit. Nucl. Instrum. Meth. A **506**, 250-303 (2003). https://doi.org/10.1016/S0168-9002(03)01368-8

[37] X. Zhou, Q. Liu, M. Wurm *et al.*, Rayleigh scattering of linear alkylbenzene in large liquid scintillator detectors. Rev. Sci. Instrum. **86**, no.7, 073310 (2015). https://doi.org/10.1063/1.4927458

[38] L. Gao, B. Yu, Y. Ding *et al.*, attenuation length measurements of a liquid scintillator with LabVIEW and a reliability evaluation of the device. Chin. Phys. C **37**, 076001 (2013). https://doi.org/10.1088/1674-1137/37/7/076001

[39] M. Wurm, F.von Feilitzsch, M. Goeger-Neff *et al.*, Optical scattering lengths in Large Liquid-Scintillator Neutrino detectors, Rev. Sci. Instrum. **81**, 053301 (2010). https://doi.org/10.1063/1.3397322

[40] Y. Zhang, Z. Y. Yu, X. Y. Li *et al.*, Complete optical model for liquid-scintillator detectors. Nucl. Instrum. Meth. A **967**, 163860 (2020). https://doi.org/10.1016/j.nima.2020.163860

[41] X. F. Ding, L. J. Wen, X. Zhou *et al.*, Measurement of fluorescence quantum yield of bis-MSB. Chin. Phys. C **39**, no.12, 126001 (2015). https://doi.org/10.1088/1674-1137/39/12/126001

[42] C. Buck, B. Gramlich, S. Wagner, Light propagation and fluorescence quantum yields in liquid scintillators. JINST **10**, no.09, P09007 (2015). https://doi.org/10.1088/1748-0221/10/09/P09007

[43] H. M. O'Keeffe, E.O'Sullivan and M. C. Chen, Scintillation decay time and pulse shape discrimination in oxygenated and deoxygenated solutions of linear alkylbenzene for the SNO+ experiments. Nucl. Instrum. Meth. A **640**, 119-122 (2011). https://doi.org/10.1016/j.nima.2011.03.027

[44] M. Yu, L. Wen, X. Zhou *et al.*, Determine Energy Nonlinearity and Resolution of $e^{\pm}$ and γ in Liquid Scintillator detectors using a universal energy response model. https://doi.org/10.48550/arXiv.2211.02467

[45] D.Adey, A. B. Balantekin, M.Bishai *et al.*, High-precision calibration of the nonlinear energy response at Daya Bay. Nucl. Instrum. Meth. A **940**, 230-242 (2019). https://doi.org/10.1016/j.nima.2019.06.031

[46] J. Dunger S. D. Biller, Multi-site Event discrimination in large liquid scintillation detectors, Nucl. Instrum. Meth. A **943**, 162420 (2019). https://doi.org/10.1016/j.nima.2019.162420

[47] A. Hocker, P. Speckmayer, J. Stelzer *et al.*, TMVA - Toolkit for multivariate data analysis, https://doi.org/10.48550/arXiv.physics/0703039.

[48] P. Speckmayer, A. Hocker, J. Stelzer *et al.*, Toolkit for multivariate data analysis, TMVA 4. J. Phys. Conf. Ser. **219**, 032057 (2010). https://doi.org/10.1088/1742-6596/219/3/032057

[49] T. Lampen, F. Garcia, A. Heikkinen *et al.*, Testing TMVA software intagging for the search of MSSM Higgs bosons at the LHC, J. Phys. Conf. Ser. **119**, 032028 (2008). https://doi.org/10.1088/1742-6596/119/3/032028

[50] L. Q. Yin, S. S. Zhang, Z. Cao *et al.* [LHAASO], expected energy spectrum of cosmic ray protons and helium below 4 PeV measured by LHAASO, Chin. Phys. C **43**, no.7, 075001 (2019). https://doi.org/10.1088/1674-1137/43/7/075001

[51] D. Guest, J. Collado, P. Baldi *et al.*, Jet Flavor Classification in High-Energy Physics with Deep Neural Networks, Phys. Rev. D **94**, no.11, 112002 (2016). https://doi.org/10.1103/PhysRevD.94.112002

[52] D. Guest, K. Cranmer and D. Whiteson, Deep Learning and its Application to LHC Physics, Ann. Rev. Nucl. Part. Sci. **68**, 161-181 (2018). https://doi.org/10.1146/annurev-nucl-101917-021019

[53] J. P. He, X. B. Tang, P. Gong, *et al.*, Spectrometry analysis based on approximation coefficients, and deep belief networks. NUCL SCI TECH 29, 69 (2018). https://doi.org/10.1007/s41365-018-0402-4

[54] X. K. Ma, H. Q. Huang, Q. C. Wang, *et al.*, Estimation of Gaussian overlapping nuclear pulse parameters based on a deep-learning LSTM model. NUCL SCI TECH 30, 171 (2019). https://doi.org/10.1007/s41365-019-0691-2

[55] Z. Qian, V. Belavin, V. Bokov *et al.*, Vertex and energy reconstruction in JUNO using machine learning

methods. Nucl. Instrum. Meth. A **1010**, 165527 (2021). https://doi.org/10.1016/j.nima.2021.165527

[56] A. Bhardwaj, J. Dutta, P. Konar *et al.*, Boosted jet techniques for a supersymmetric scenario with gravitino LSP. JHEP **10**, 083 (2020). https://doi.org/10.1007/JHEP10(2020)083

[57] Y. Z. Li, Z. Qian, J. H. He *et al.*, Improvement of machine-learning-based vertex reconstruction for large liquid scintillator detectors with multiple types of PMTs. NUCL SCI TECH 33, 93 (2022). https://doi.org/10.1007/s41365-022-01078-y

[58] H. L. Liu, H. B. Ji, J. M. Zhang *et al.*, A Novel Approach for Feature Extraction from a Gamma Ray Energy Spectrum Based on Image Descriptor Transfer for radionuclide Identification. NUCL SCI TECH 33, 158 (2022). https://doi.org/10.1007/s41365-022-01150-7

[59] C. F. Yang, Y. B. Huang, J. L. Xu *et al.*, Reconstruction of a muon bundle in the JUNO central detector. Nucl. Sci. Tech. **33**, no.5, 59 (2022). https://doi.org/10.1007/s41365-022-01049-3